



Making the Required Move to External Solr

Mark Cooper, Brian Hoffman, Christine Di Bella

November 3, 2021





What we'll cover

- What is Solr?
- Solr in ArchivesSpace
- Why ArchivesSpace will need external Solr
- Installing and configuring Solr
- Resources
- Q&A





What is Solr?

- Open source search platform built on Apache Lucene
- Versatile and scalable
- Large, active community
- Used by many open source and proprietary sites and projects, including Blacklight and ArcLight, the Internet Archive, JSTOR, and Netflix





Solr in ArchivesSpace

- It's the search engine for ArchivesSpace
- When we say "index" this is the part we're talking about
- Determines weighting in search results
- Through the current version, embedded into releases of the application, though it was also possible to use external Solr
- "External Solr" means ArchivesSpace points to an instance of Solr that lives outside of your ArchivesSpace instance
- Similar to how MySQL needs to be installed separately





Why ArchivesSpace will need external Solr

- ArchivesSpace currently uses Solr v4.10.4 (released in 2014)
- Solr itself is currently on v8.10.1
- Because of changes made by the Solr maintainers, Solr is no longer embeddable
- All versions of Solr past the one we're using require external Solr (For a full technical explanation see: <https://cwiki.apache.org/confluence/display/solr/WhyNoWar>)





Why ArchivesSpace will need external Solr

- Staying on this old version has risk since it is not supported
- Staying on this old version limits our ability to improve search capabilities in the application
- We stayed on the older version as long as we could to minimize inconvenience to the community, but we need to move on.
- **Please note:** if you use a hosting provider to manage your ArchivesSpace, your hosting provider will manage Solr too





Benefits of managing your own Solr

- Ability to monitor what's going on with your indexing
- Ability to use any of the features that are in the Solr documentation to manage the server
- Would allow people to use Solr cloud configurations; potentially easier integration with other applications, better performance





Different ways to run external Solr

- Install locally
- Cloud
- Ansible
- Use a Docker container

Some ArchivesSpace-specific Solr files will be provided in an ArchivesSpace release.





Overview of steps to run ArchivesSpace with external Solr

(This will be part of your upgrade process once an ArchivesSpace release with the Solr update comes out)

1. Stop ArchivesSpace
2. Install Solr using your preferred method
3. Start Solr





Overview of steps to run ArchivesSpace with external Solr

4. Apply ArchivesSpace configuration using the files and directions provided with the release
5. Ensure Solr URL is correct in the config file
6. Reindex ArchivesSpace using your preferred method (easiest is to delete `indexer_state` files)
7. Restart ArchivesSpace



8. Success:

Nothing changes for you and your users from this point on and everyone goes on with their day



Resources for downloading and installing Solr

- General directions from Solr:
https://solr.apache.org/guide/8_9/installing-solr.html
- Technical documentation from ArchivesSpace:
<https://archivesspace.github.io/tech-docs/provisioning/solr.html>
- Tutorial to come
- ArchivesSpace members are eligible for technical support to get help setting up Solr:
<https://archivesspace.org/member-area/technical-support>
- More to come as the release approaches





Q&A

Feel free to ask anything Solr-related now, or follow up with us at ArchivesSpaceHome@lyrasis.org after the webinar.

